

Möglichkeiten und Grenzen des internationalen Vergleichs fremdenfeindlicher Vorurteile

Weins, Cornelia

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Weins, C. (2008). Möglichkeiten und Grenzen des internationalen Vergleichs fremdenfeindlicher Vorurteile. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, Methoden und Instrumente der Sozialwissenschaften 2008/1, 25-43. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205158>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Möglichkeiten und Grenzen des internationalen Vergleichs fremdenfeindlicher Vorurteile¹

Cornelia Weins

Zusammenfassung

In wissenschaftlichen Analysen werden regelmäßig vergleichende Aussagen über das Ausmaß von Vorurteilen in Staaten getroffen, ohne angemessen zu prüfen, ob die verwendeten Instrumente in den untersuchten Staaten auch dasselbe messen. Dies ist erstaunlich, ist doch das Ausmaß von Vorurteilen in einer Gesellschaft von unmittelbar politischer und wissenschaftlicher Brisanz. Mit Daten des Eurobarometers 53 wird in diesem Beitrag daher die Prüfung der Messinvarianz dichotomer Vorurteilsindikatoren demonstriert. Eine bisher in der Literatur nicht ausreichend beachtete Ursache für die fehlende internationale Vergleichbarkeit von Vorurteilsskalen ist *item-nonresponse*. Die Verzerrungen durch item-nonresponse können durch den Einsatz theoretisch fundierter Methoden behoben oder zumindest eingrenzt werden, wie in diesem Beitrag gezeigt wird.

Problemaufriss: Fragestellungen vergleichender Untersuchungen

Seit den achtziger Jahren haben vergleichenden Analysen fremdenfeindlicher Vorurteile sprunghaft zugenommen. Dies hängt wesentlich mit der Verfügbarkeit von Sekundärdaten zusammen. Im Auftrag der Europäischen Kommission wurde erstmals im Eurobarometer 30 (1988) der Themenschwerpunkt „Einstellungen gegenüber Minderheiten“ implementiert. Seitdem werden Indikatoren zur Messung von Vorurteilen gegenüber Minderheiten in den Eurobarometern regelmäßig erhoben. Vergleichend können Vorurteile auch mit den Daten des ISSP 1995 und 2003 sowie seit einigen Jahren mit den Daten des European Social Survey analysiert werden.

Von ihrer Fragestellung lassen sich vergleichende Analysen fremdenfeindlicher Vorurteile grob in zwei Kategorien einteilen: Erstens gibt es Studien, die primär Unterschiede oder Gemeinsamkeiten in der Entstehung fremdenfeindlicher Vorurteile in verschiedenen Staaten untersuchen. Staaten werden hier getrennt analysiert und die Analyseergebnisse anschließend miteinander verglichen. Exemplarisch für diese Art von Studien ist die umfangreiche Arbeit von Zick (1997), der eine Substichprobe des Eurobarometers 30 analysiert hat. Eine geringe formale Bildung hat sich beispielsweise als ein interkulturell stabiler Prädiktor von Vorurteilen erwiesen, während Nationalstolz in einigen aber nicht allen europäischen Staaten die Entstehung von Vorurteilen begünstigt (vgl. auch Westle 1999, Weins 2004). Das primäre Ziel dieser Studien besteht darin herauszufinden, ob Zusammenhänge, die in einem Staat bestehen, auch in anderen Staaten beobachtet werden können oder nicht. Unterschiede in der Entstehung von Vorurteilen in verschiedenen Staaten werden im Anschluss an die Analyse – also ex-post – interpretiert. Man kann diese Studien aus vergleichender Perspektive als *strukturorientiert* bezeichnen, weil ein Vergleich auf dieser Ebene voraussetzt, dass ein „instrument measures the same construct across cultures“ (van de Vijver 2003, 143 f.).

1 Für wertvolle Hinweise danke ich Jost Reinecke.

Darüber hinausgehend wurde - zweitens - versucht, Unterschiede im Ausmaß der Vorurteile zwischen verschiedenen Staaten (oder besser: zwischen Befragten verschiedener Staaten) zu erklären. Dies geschieht, indem Merkmale auf Ebene der Staaten berücksichtigt werden, von denen ausgegangen wird, dass sie die Entstehung fremdenfeindlicher Vorurteile in einem Staat begünstigen. Aus vergleichender Perspektive sind Studien, die Merkmale von Staaten in die Erklärung einbeziehen, interessanter als getrennte Analysen für einzelne Staaten, weil die Merkmale, die als relevant für die Entstehung von Gruppenunterschieden angesehen werden, direkt modelliert und empirisch geprüft werden können. Einige Autoren sind der Ansicht, dass „circumstances, variables or processes“ zu identifizieren „which describe what is about these groups, cultures or nations that affect the racist phenomena one is interested in“ die eigentliche Aufgabe vergleichender Analysen fremdenfeindlicher Vorurteile sei (Ter Wal und Verkuyten 2000, 4). Als relevante Prädiktoren fremdenfeindlicher Vorurteile auf Ebene der Staaten werden insbesondere die *Höhe der Zuwanderung* und die *makro-ökonomischen Bedingungen* angesehen (Quillian 1995, Kunovich 2001, Coenders 2001, Scheepers et al. 2002, Semyonov et al. 2006). Diese Studien orientieren sich theoretisch an dem in der *American Sociological Review* erschienenen Beitrag von Quillian (1995), der der These zum Einfluss der Größe der zugewanderten Minderheit für die Entstehung von Vorurteilen in wissenschaftlichen Analysen zu neuer Prominenz verholfen hat. Quillians These ist, dass die Mehrheitsangehörigen ihre eigene Position mit zunehmender Größe der Minderheit und bei schlechten makro-ökonomischen Bedingungen in Gefahr sehen (Gruppenbedrohung) und mit Vorurteilen gegenüber Minderheitenangehörigen reagieren. Anders ausgedrückt: je höher die Zuwanderung und je schlechter die wirtschaftliche Lage in einem Staat, umso höher ist das Ausmaß der Vorurteile. Zur Begründung zieht Quillian zwei Hypothesen Hubert M. Blalocks (1967) zum Einfluss der Größe der Minderheit (*competition* und *power-threat*) auf Diskriminierung heran. Die theoretische Begründung des Zusammenhangs zwischen dem Ausländeranteil und dem Ausmaß fremdenfeindlicher Vorurteile in den europäischen Staaten ist nicht Gegenstand dieses Beitrages. Hingewiesen sei lediglich darauf, dass von allen genannten Autoren vernachlässigt wird, dass mit zunehmender Größe der Minderheit auch die Kontakte zwischen Minderheits- und Mehrheitsangehörigen zunehmen und die empirischen Evidenzen mehrheitlich dafür sprechen, dass Kontakte Vorurteile abbauen (Weins 2004, 46-50). Quillian gelangt zu dem beeindruckenden Ergebnis, dass 70% der Unterschiede zwischen Staaten durch den (Nicht-EG)-Ausländeranteil und die makro-ökonomische Lage der Staaten erklärt werden können (Quillian 1995, 603). Die These zunehmender Vorurteile mit der Größe der Minderheit und schlechten makro-ökonomischen Bedingungen wird von allen genannten Autoren (in modifizierter Form) mit Hilfe von Mehrebenenanalysen untersucht, d. h. es werden gleichzeitig Prädiktoren auf Ebene der Staaten und der individuellen Ebene berücksichtigt. Mit Ausnahme von Semyonov et al. (2006) wird nur ein Zeitpunkt betrachtet. Semyonov et al. (2006) vergleichen das Ausmaß der Vorurteile in den 12 EG-Staaten für vier verschiedene Zeitpunkte (1988, 1994, 1997 und 2000) mit Daten des Eurobarometers, d. h. die Vergleichsperspektive beinhaltet auch die Entwicklung des Ausmaßes der Vorurteile in den EG-Staaten über die Zeit (vgl. auch Wilkes, Guppy und Farris 2007). Weil die Skalenwerte von Befragten verschiedener Staaten direkt miteinander verglichen werden, bezeichnet van de Vijver (2003) diese Studien als *niveaurorientiert*.²

In Abbildung 1 ist das durchschnittliche Ausmaß der Vorurteile in den 12 EG-Staaten abgebildet, wie sie Semyonov et al. (2006) für das Jahr 2000 mit dem Eurobarometer auf Basis von vier Indikatoren ermittelt haben. Man sieht deutliche Unterschiede zwischen Staaten: In Griechenland und Bel-

2 Niveaurorientiert sind natürlich auch Studien, in denen Mittelwerte für Staaten zwar verglichen, nicht aber erklärt werden.

gien sind die Skalenwerte am höchsten, in Spanien und Luxemburg am geringsten. Diese Unterschiede zwischen Staaten werden von den Autoren als Unterschiede in der Ausprägung der Vorurteile interpretiert.

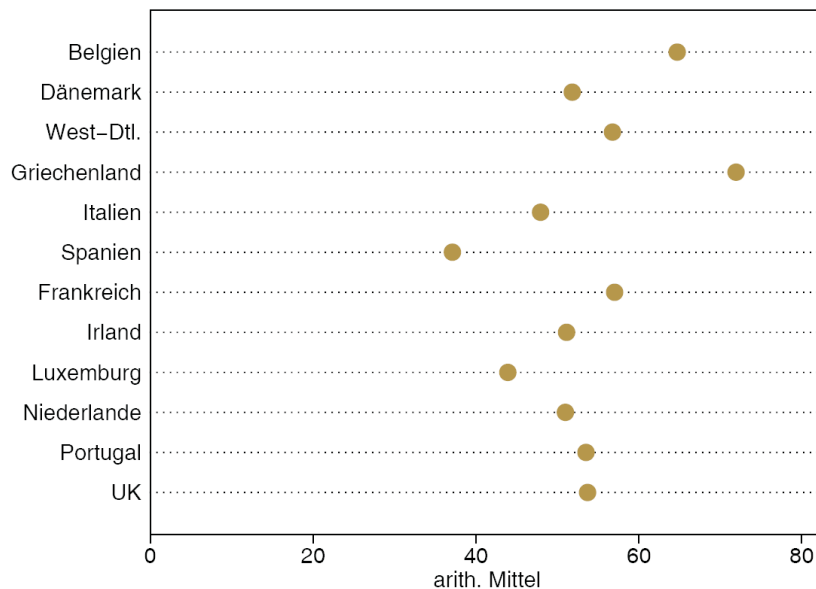


Abbildung 1: Anti-foreigner Sentiment in Europe

Quelle: Semyonov et al. (2006), S. 433/436, (eigene Berechnung, ungewichtet)

Erstaunlicherweise wird nicht untersucht, ob die beobachteten Unterschiede zwischen Staaten tatsächlich als Unterschiede in den Vorurteilen interpretiert werden können. Semyonov et al. (2006) demonstrieren lediglich die Eindimensionalität der für die Skalenbildung verwendeten Indikatoren über eine explorative Faktorenanalyse, in die alle Befragten eingehen. Es bleibt damit offen, ob die beobachteten Mittelwertunterschiede als Unterschiede in den Vorurteilen interpretiert werden können.

Einerseits sind die statistischen Verfahren zur Prüfung der Bedeutung der Größe der Minderheit und der makro-ökonomischen Bedingungen für die Genese fremdenfeindlicher Vorurteile elaboriert (Mehrebenenanalysen). Andererseits wird nicht einmal ansatzweise geprüft, ob die Indikatoren in verschiedenen Staaten dasselbe messen und ob eine Invarianz der Messinstrumente etabliert werden kann, die den Vergleich des Niveaus der Vorurteile erlaubt. Für strukturorientierte Analysen sind die Anforderungen an die Messinvarianz geringer als für niveaurorientierte Analysen, weil die Skalenwerte von Befragten unterschiedlicher Staaten nicht direkt miteinander verglichen werden. Im folgenden wird mit den Indikatoren von Semyonov et al. (2006) die Prüfung der Messinvarianz demonstriert. Beachtet werden muss, dass die im Eurobarometer verwendeten Indikatoren zur Messung von Vorurteilen dichotom sind und sehr hohe Anteile fehlender Werte aufweisen.

Indikatoren zur Messung von Vorurteilen

Semyonov et al. (2006) verwenden vier Indikatoren zur Messung von Vorurteilen. Die ersten drei Indikatoren entstammen einer Itematterie. Hier wird erfasst, ob die Befragten der Ansicht sind, dass Angehörige rassistischer, religiöser oder kultureller Minderheiten *das Sozialsystem ausnutzen*, die *Unsicherheit* oder *Arbeitslosigkeit* im eigenen Staat erhöhen. Auf diese Aussagen konnten die Befragten zustimmend, ablehnend oder mit weiß nicht antworten. Als vierter Indikator wurde die Frage aufgenommen, ob es nach Ansicht der Befragten „*nicht viele*“, „*viele, aber nicht zu viele*“ oder „*zu viele*“ Angehörige religiöser, rassistischer oder kultureller Minderheiten im eigenen Staat gibt.³ Auch hier stand den Befragten als Option die Kategorie weiß nicht zur Verfügung. Aufgrund der schiefen Verteilung des letztgenannten Indikators wurde dieser dichotomisiert, wobei die Kategorien viele, aber nicht zu viele und nicht zu viele zusammengefasst wurden.

Tabelle 1: Vorurteilsindikatoren (in Klammern: Kurzbezeichnung)

| |
|---|
| For each of the following opinions, please tell me whether you tend to agree or tend to disagree? |
| 1. People from these minority groups* abuse the system of social benefits (<i>Sozialsystem</i>) |
| 2. The presence of people from these minority groups* is a cause of insecurity (<i>Unsicherheit</i>) |
| 3. The presence of people from these minority groups* increases unemployment in (COUNTRY) (<i>Arbeitslosigkeit</i>) |
| 4. Again, speaking generally about people from minority groups in terms of race, religion and culture, do you think there are not many, a lot but not too many, or too many of them living in (OUR COUNTRY) (<i>Zahl</i>) |
| *race, religion and culture |
| Eurobarometer 53.0, 2000 |

Bei rassistischen, religiösen und kulturellen Minderheiten kann es sich sowohl um primäre (angestammte) als auch um sekundäre (zugewanderte) Minderheiten handeln (Francis 1965). Dies impliziert, dass die Referenzgruppe von der Größe der angestammten bzw. zugewanderten Minderheit abhängt. Die Indikatoren beziehen sich daher nicht automatisch auf Zuwanderer, auch wenn dies in den Staaten, die nach dem zweiten Weltkrieg eine hohe Zuwanderung verzeichnet haben, vermutet werden kann (vgl. zu primären Minderheiten in Europa Pan und Pfeil 2000). An dieser Stelle kann nicht geprüft werden, inwieweit die in andere Sprachen übersetzten Items von der englischen Version abweichen. Übersetzungsprobleme oder -fehler können eine Ursache der Nicht-Vergleichbarkeit eines

3 Die Frageformulierung ist nicht für alle vier Zeitpunkte, die Semyonov et al. (2006) betrachten, identisch (!). Die Autoren geben aber lediglich die Formulierungen für 1988 an (Semyonov et al. 2006, 446). Die wichtigsten Unterschiede sind die, dass 1988 die Zielgruppe „foreigners“ und nicht religiöse, rassistische oder kulturelle Minderheiten als Zielgruppe genannt wurden, daß 1988 „exploit our social welfare system“ (nicht: abuse the system of social benefits) und „is a cause of crime and violence“ (nicht: is a cause of insecurity) vorgegeben wurden. Weil im vorliegenden Beitrag der Ländervergleich für das Jahr 2000 untersucht wird und nicht die Entwicklung der Vorurteile zwischen 1988 und 2000, wird auf Veränderungen in den Frageformulierungen und deren Konsequenzen für die Vergleichbarkeit des Messinstrumentes über die Zeit nicht eingegangen. Hingewiesen sei aber darauf, dass die Invarianz der Messinstrumente über die Zeit ebenfalls mit den in diesem Beitrag dargestellten Methoden erfolgen kann.

Instrumente zwischen Staaten sein (vgl. für die Messung von Geschlechterrollen mit dem ISSP Braun 2006, 146 f.).

Bei diesen Itembatterien treten im Eurobarometer (nicht nur im Jahr 2000) sehr hohe Anteile fehlender Werte auf. In Tabelle 2 ist abgetragen, wie viel Prozent der Befragten in einem Staat⁴ keinen fehlenden Wert haben (Spalte 0), einen fehlenden Wert (Spalte 1) usw. Wie man sieht, variiert der Anteil fehlender Werte deutlich zwischen den einzelnen Staaten. In Belgien haben 20% der Befragten einen oder mehrere fehlende Werte, in Großbritannien sind es schon 40%. Die geringsten Anteile fehlender Werte treten bei der Frage zur Zahl der Immigranten auf. In Griechenland resultieren die fehlenden Werte vor allem aus der Nicht-Beantwortung der Aussage, Minderheiten nützten das Sozialsystem aus.

Tabelle 2: Anteile fehlender Werte

| | Zahl fehlender Werte* | | | | | Total | (N) |
|-----------------|-----------------------|----|----|---|---|-------|--------|
| | 0 | 1 | 2 | 3 | 4 | | |
| Belgien | 81 | 11 | 5 | 2 | 1 | 100 | (1063) |
| Dänemark | 75 | 18 | 4 | 2 | 0 | 100 | (1000) |
| Westdeutschland | 62 | 21 | 10 | 5 | 1 | 100 | (1015) |
| Griechenland | 77 | 17 | 4 | 2 | 0 | 100 | (1004) |
| Italien | 66 | 21 | 9 | 4 | 1 | 100 | (1000) |
| Spanien | 71 | 17 | 7 | 4 | 1 | 100 | (1000) |
| Frankreich | 78 | 13 | 6 | 1 | 1 | 100 | (1002) |
| Irland | 65 | 19 | 9 | 5 | 2 | 100 | (1000) |
| Luxemburg | 71 | 18 | 7 | 4 | 1 | 100 | (600) |
| Niederlande | 72 | 18 | 7 | 2 | 1 | 100 | (975) |
| Portugal | 63 | 21 | 9 | 6 | 2 | 100 | (1000) |
| Großbritannien | 60 | 23 | 10 | 5 | 3 | 100 | (1070) |

*Zeilenprozente

Das Standardverfahren zum Umgang mit fehlenden Werten besteht nach wie vor darin, alle Fälle mit mindestens einem fehlenden Wert aus der Analyse auszuschließen (listenweiser Fallausschluss).⁵ Der listenweise Fallausschluss ist allerdings nur dann gerechtfertigt, wenn die kompletten Fälle - hier

4 Im folgenden wird ausschließlich Großbritannien ohne Nordirland untersucht. Für Nordirland wurde die Messinvarianz aufgrund kleiner Fallzahlen nicht untersucht. Weniger als 3% der über 15jährigen Bevölkerung des Vereinigten Königreichs (Grundgesamtheit des ISSP) haben im Jahr 2000 in Nordirland gelebt (INRA 2000).

5 Semyonov et al. (2006) versuchen der drastischen Reduktion der Fallzahlen entgegenzuwirken, indem sie den Index als Anteil der Zustimmungen an den gültigen Antworten berechnen. Diese in den Sozialwissenschaften häufig eingesetzte Methode der Behandlung fehlender Werte bei der Skalenbildung hat den Nachteil, dass nicht angegeben werden kann, unter welchen Bedingungen sie unverzerrte Ergebnisse liefert, die Varianzen tendenziell überschätzt werden und die Skalenwerte der Personen auf unterschiedlichen Items beruhen (vgl. Schafer & Graham 2002, 157 f.). Bei vergleichenden Analysen entsteht eine zusätzliche Verzerrung dadurch, dass dieselben Items in Abhängigkeit von der Höhe des item-nonresponse in verschiedenen Staaten eine unterschiedliche Bedeutung für die Skalen haben.

also zwischen 60% (Großbritannien) und 80% (Belgien) der Befragten - eine einfache Zufallsstichprobe aus allen Fällen darstellen. Der Ausfallmechanismus ist dann komplett zufällig, d. h. *missig completely at random* (MCAR). Dies ist in der Regel eine unrealistische Annahme: Unterscheiden sich Befragte mit fehlenden Werten in ihrem Antwortverhalten von Befragten mit vollständigen Angaben, dann führt eine Analyse der kompletten Fälle zu verzerrten inhaltlichen Ergebnissen. Beispielsweise stimmen in Frankreich Personen mit unvollständigen Angaben in signifikant geringerem Ausmaß den Vorurteilsindikatoren zu, für die Angaben vorhanden sind. Durch listenweisen Fallausschluss würde das Niveau der Vorurteile in Frankreich daher überschätzt. Ursächlich dafür ist, dass in Frankreich höher Gebildete eher Angaben verweigern und Vorurteile bei höher Gebildeten geringer ausgeprägt sind. Das Niveau der Vorurteile in Frankreich würde falsch eingeschätzt und dadurch Unterschiede im Niveau der Vorurteile zwischen Staaten, weil nicht in allen Staaten Befragte mit vollständigen Angaben (signifikant) stärker den Vorurteilsindikatoren zustimmen. Auch Aussagen über individuelle Determinanten fremdenfeindlicher Vorurteile werden durch item-nonresponse verzerrt, wenn die Struktur der Beziehungen zwischen Prädiktoren und Vorurteilsindikatoren bei Befragten mit und ohne fehlende Werte unterschiedlich ist: Mit steigendem Bildungsniveau sinkt die Zustimmung zu den Vorurteilsindikatoren beispielsweise bei französischen Befragten mit fehlenden Werten nicht so stark wie bei französischen Befragten mit kompletten Angaben. Listenweiser Fallausschluss führt demnach zu einer Überschätzung der Bedeutung des Bildungsniveaus für Vorurteile in Frankreich. Die Auswirkungen von *item-nonresponse* für die internationale Vergleichbarkeit von Messinstrumenten wurden m. W. bisher nicht ausreichend thematisiert (vgl. aber zu *unit-nonresponse* Couper und de Leeuw 2003).

Ein listenweiser Ausschluss fehlender Werte ist zudem nicht sehr effizient, weil die beobachteten Informationen von Personen, für die (mindestens) eine Angabe fehlt, unberücksichtigt bleiben. Bei den Befragten, die lediglich eine Aussage nicht beantwortet haben (Spalte „1“ in Tabelle 32), liegen drei Informationen zu den Vorurteilsindikatoren vor, eine Angabe fehlt. Der listenweise Fallausschluss führt dazu, dass die drei vorhandenen Informationen unberücksichtigt bleiben („weggeworfen“ werden). Das heißt, selbst wenn der Ausfallmechanismus komplett zufällig sein sollte, werden die beobachteten Informationen nicht angemessen ausgeschöpft. In Großbritannien führt der listenweise Fallausschluss allein durch die Indikatoren zur Messung von Vorurteilen zu einer drastischen Reduktion der Stichprobengröße auf 637 Befragte. Differenzierte Analysen für Subgruppen werden erschwert und die Unsicherheit über die Parameter der Grundgesamtheit (Standardfehler) größer.

Mehrfache Ergänzung der fehlenden Werte

Aus diesen Gründen ist es sinnvoll, alle beobachteten Werte für die Analyse zu nutzen. Dies geschieht hier, indem für jeden fehlenden Wert mehrere (multiple) „plausible“ Werte ergänzt (imputiert) werden (Little & Rubin 2002, einführend Allison 2002). Wichtig ist, dass es sich bei diesen „plausiblen“ Werten nicht um beliebig oder willkürlich (ad hoc) eingesetzte Werte handelt.⁶ Die Ergänzung der fehlenden Werte erfolgt vielmehr unter Ausnutzung aller beobachteten Informationen.

6 Eine ad hoc-Methode wäre beispielsweise die Ersetzung eines fehlenden Wertes durch den arithmetischen Mittelwert des betreffenden Indikators. Die Ersetzung durch den Mittelwert führt zu verzerrten Varianzen, Kovarianzen und Korrelationen und ist daher als Ersetzungsmethode ungeeignet. Vgl. einführend in verschiedene Methoden der Behandlung fehlender Werte sowie deren Vor- und Nachteile Schafer und Graham (2002).

Berücksichtigt werden bei der Imputation (Ergänzung) alle Variablen, die für die substanzwissenschaftliche Analyse von Interesse sind. Zudem kann die Qualität der ergänzten Werte verbessert werden, indem zusätzlich Variablen berücksichtigt werden, die einen engen Zusammenhang mit Variablen mit hohen Anteilen fehlender Werte aufweisen (vgl. zu Interviewermerkmalen z. B. Schröpfer 2004).

Zwei Punkte müssen beachtet werden, damit es sich um *korrekte Imputationen* nach Rubin handelt: erstens muss stochastisch imputiert werden und zweitens muss berücksichtigt werden, dass die der Imputation zugrundeliegenden Parameter stichprobenabhängig variieren. Diese beiden Punkte lassen sich am einfachsten an zwei normalverteilten Variablen X und Y veranschaulichen. X sei vollständig beobachtet und in Y treten fehlende Werte auf. Die Ergänzung der fehlenden Werte könnte so erfolgen, dass für die vollständigen Fälle (in Abbildung 2 (a) durch dunkle Punkte dargestellt) eine Regressionsgleichung $\hat{Y} = a + b X$ geschätzt würde. Für die Fälle mit fehlenden Werten in X könnte dann der Schätzwert der Regressionsgeraden, \hat{Y} , als Wert ergänzt werden, wie es in Abbildung 3 (a) durch die hellen Punkte dargestellt ist. Bei einem solchen Vorgehen würde jedoch die Korrelation zwischen der vervollständigten Y-Variablen und der X-Variable überschätzt, weil alle ergänzten Werte exakt auf der Regressionsgeraden lägen.

Um die Streuung um die Regressionsgerade zu reproduzieren, wird deshalb der konditionale Mittelwert \hat{Y} plus eine Zufallskomponente (ein zufällig gezogener Wert aus der residualen Normalverteilung) ergänzt, vgl. Abbildung 2 (b). Unvollständige Fälle mit derselben Ausprägung in X erhalten dadurch unterschiedliche Werte für Y.

Werden die Regressionsparameter a und b aus der Stichprobe geschätzt, wie gerade skizziert, so wird vernachlässigt, dass diese stichprobenabhängig von den Parametern der Grundgesamtheit - beispielhaft dargestellt durch die gestrichelte Linie in Abbildung 2 (c) - abweichen. Um die Unsicherheit über die Parameter der Grundgesamtheit (hier α und β) zu berücksichtigen, werden daher für jeden fehlenden Wert *mehrere Werte* ergänzt, und zwar jeweils auf Basis unterschiedlicher Werte für die Regressionsparameter a und b. Bei diesen handelt es sich um Zufallsziehungen aus ihren a-posteriori-Verteilungen. Die für einen fehlenden Wert ergänzten Werte unterscheiden sich daher aufgrund der Zufallskomponente und der Variation in den zugrundeliegenden Regressionsparametern. In der Regel werden zwischen 5 und 10 Werte für jeden fehlenden Wert imputiert (Little und Rubin 2002, 87).

Weil mehrere Werte ergänzt werden, liegen auch mehrere vervollständigte Datensätze vor, bei 5 ergänzten Werten 5 vervollständigte Datensätze. Diese Datensätze müssen zunächst getrennt analysiert werden. Anschließend können die Ergebnisse nach einfachen Regeln (Little und Rubin 2002, 86 f.) zu einer gemeinsamen Schätzung integriert werden. Die Unterschiede in den Analyseergebnissen der vervollständigten Datensätze werden dabei zur Korrektur der Standardfehler in den Einzelanalysen genutzt.⁷

7 Die in den Einzelanalysen geschätzten Standardfehler unterschätzen den tatsächlichen Standardfehler, weil imputierte Werte wie beobachtete Werte behandelt werden und dadurch die Stichprobengröße überschätzt wird.

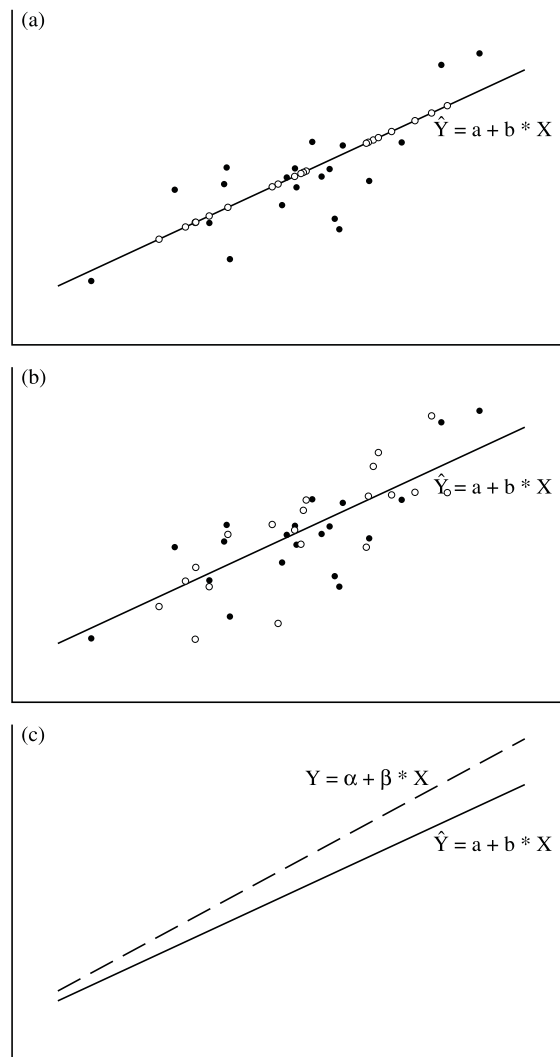


Abbildung 2: Imputation fehlender Werte

Im Gegensatz zum listenweisen Fallausschluss wird mit einer multiplen Imputation die durch die beobachteten Daten erklärbare Verzerrung durch fehlende Werte beseitigt. Der Datenausfall muss daher nicht komplett zufällig sein. Eine multiple Imputation führt zu korrekten Schätzungen auch dann, wenn die Ausfälle von beobachteten (d. h. im Imputationsmodell enthaltenen) Werten abhängen, also beispielsweise Befragte mit höherer Schulbildung eher die Angabe zu den Vorurteilsindikatoren verweigern, wie für Frankreich beobachtet wurde (*missing at random*, MAR).⁸ Und zwar deshalb,

⁸ Eine multiple Imputation setzt genau genommen einen ignorierbaren Ausfallmechanismus voraus. Der Ausfallmechanismus ist ignorierbar, wenn der Datenausfall MAR ist und sich die Parameter des den Datenausfall erzeugenden Prozesses und der kompletten Daten nicht gegenseitig beeinflussen. MAR ist die wich-

weil die Information zur Schulbildung zur Schätzung der plausiblen Werte genutzt wird. Eine multiple Imputation setzt aber voraus, dass der Datenausfall nicht von den fehlenden Werten selbst abhängt (*not missing at random*, NMAR)⁹, was in der Regel schwierig zu begründen ist. In jedem Fall ist eine multiple Imputation wegen der geringeren Anforderungen an den Datenausfall und der besseren Nutzung der vorhandenen Informationen dem listenweisen Fallausschluß vorzuziehen.

Bei einer komparativen Analyse muss die Imputation getrennt für jede Gruppe (Staat) erfolgen, um die Struktur der Beziehungen zwischen den Variablen innerhalb der einzelnen Staaten zu berücksichtigen. Ins Imputationsmodell wurden die Vorurteilsindikatoren aufgenommen und zusätzlich alle Prädiktoren die Semyonov et al. (2006) auf der individuellen Ebene berücksichtigt haben.¹⁰ Die Imputation wurde mit Stata 10 durchgeführt (Royston 2005, 2004). Für jeden fehlenden Wert werden fünf Werte ergänzt, d. h. es liegen (12x)5 vervollständigte Datensätze vor. Dargestellt werden im folgenden die Ergebnisse über alle imputierten Datensätze. Tabelle 3 enthält die Zustimmung zu den Vorurteilsindikatoren auf Basis der imputierten Datensätze.

Tabelle 3: Zustimmung zu den Vorurteilsindikatoren (imputiert)

| | Sozialsystem | Unsicherheit | Arbeitslosigkeit | Zahl (zu viele) | N |
|------------------|--------------|--------------|------------------|-----------------|------|
| Belgien | 71 | 61 | 71 | 56 | 1063 |
| Dänemark | 59 | 65 | 53 | 35 | 1000 |
| Westdeutschland | 64 | 54 | 66 | 47 | 1015 |
| Griechenland | 59 | 80 | 88 | 58 | 1004 |
| Italien | 49 | 47 | 51 | 45 | 1000 |
| Spanien | 43 | 36 | 45 | 23 | 1000 |
| Frankreich | 70 | 54 | 58 | 43 | 1002 |
| Irland | 69 | 47 | 53 | 36 | 1000 |
| Luxemburg | 51 | 43 | 53 | 26 | 600 |
| Niederlande | 58 | 51 | 56 | 42 | 975 |
| Portugal | 65 | 54 | 68 | 38 | 1000 |
| Großbritannien | 68 | 58 | 39 | 47 | 1070 |
| $\bar{X}_{agg.}$ | 60 | 60 | 52 | 41 | |

Gemeinsame Schätzung über die 5 vervollständigten Datensätze.

N=Fallzahl pro Datensatz, $\bar{X}_{agg.}$ = arithmetisches Mittel über die Anteile der Staaten

tigere der beiden Bedingungen (Little und Rubin 2002, 119), weshalb die Anforderung häufig auf MAR reduziert wird.

- 9 Das klassische Beispiel für einen nicht-zufälligen Ausfallmechanismus sind Einkommen von Frauen. Das Arbeitsangebot von Frauen hängt von der Höhe des (erwarteten) Einkommens selbst ab. Fehlende Einkommensangaben für nicht-erwerbstätige Frauen sind daher *not missing at random*. Hier muss der Ausfallmechanismus direkt modelliert werden, wie es klassischerweise durch Heckmans Selektionsmodell geschieht (Engelhardt 1999).
- 10 Das sind Schulbildung (Alter bei Beendigung der Ausbildung), Geschlecht, Alter (in Jahren), Ehestatus (verheiratet=1), Links-Rechts-Selbsteinstufung, Einkommen (im niedrigsten Quartil=1), arbeitslos (vgl. zur Kodierung Semyonov et al. 2006, 446).

Man sieht, dass die Zustimmung zu den Indikatoren deutlich zwischen Staaten und zwischen Indikatoren für einzelne Staaten variiert. Die für vergleichende Analysen zentrale Frage ist nun, ob und auf welchem Niveau die vier dichotomen Indikatoren in den verschiedenen Staaten dasselbe messen. Die Prüfung der Messinvarianz erfolgt mit multiplen Gruppenvergleichen konfirmatorischer Faktorenanalysen (vgl. Bollen 1989, Reinecke 2005).¹¹

Multiple Gruppenvergleiche konfirmatorischer Faktorenmodelle

Faktorenanalysen werden bei metrischen Indikatoren auf Basis von Kovarianz- oder Korrelationsmatrizen (Pearsons r) berechnet. Bei dichotomen Indikatoren ist dies nicht möglich, weil kein Intervallskalenniveau unterstellt werden kann. Ausgangspunkt der faktorenanalytischen Prüfung sind hier tetrachorische Korrelationen (Olsson 1979). Die Berechnung tetrachorischer Korrelationen beruht auf der Annahme, dass es sich bei den beobachteten, dichotomen Merkmalen X um grobe Messung eigentlich kontinuierlicher Eigenschaften X^* handelt, also um künstliche Dichotomien. Tetrachorische Korrelationen geben die Korrelationen zwischen den zugrundeliegenden, kontinuierlichen Variablen X^* an, nicht die zwischen den beobachteten Indikatoren. Eine dichotome Variable X nimmt den Wert eins an (hier: ein Befragter stimmt der Aussage zu bzw. gibt beim Indikator Zahl „zu viele“ an), wenn auf der zugrundeliegenden Variablen X^* ein bestimmter Schwellenwert v überschritten wird.

In einem konfirmatorischen Faktorenmodell wird das Messmodell, das die Beziehungen zwischen Indikatoren und Faktoren beinhaltet, vorgegeben. Anhand verschiedener Statistiken (wie χ^2 , RMSEA, vgl. Reinecke 2005) lässt sich dann beurteilen, ob das theoretisch postulierte Modell angemessen ist oder nicht. Im vorliegenden Beispiel muss ein einfaktorielles Modell vorgegeben werden, weil alle vier Indikatoren ein einziges latentes Konstrukt - nämlich Vorurteile - messen sollen. Bei *multiplen Gruppenvergleichen* werden die konfirmatorischen Faktorenanalysen gleichzeitig für mehrere Gruppen - hier 12 Staaten - berechnet. Multiple Gruppenvergleiche erlauben die Prüfung der Messinvarianz der Indikatoren, weil Parameter (z. B. Faktorenladungen) über die verschiedenen Gruppen gleichgesetzt werden können. Multiple Gruppenvergleiche erlauben die Schätzung von Unterschieden zwischen Gruppen auf der latenten Ebene, d. h. hier im Ausmaß der Vorurteile zwischen Staaten¹². Inhaltlich sinnvoll interpretierbar sind solche Unterschiede aber nur unter bestimmten Voraussetzungen.

Messinvarianz der Indikatoren

Hängt die Wahrscheinlichkeit der Zustimmung bzw. Ablehnung der vier Indikatoren ausschließlich von den Vorurteilen der befragten Personen ab, nicht jedoch von ihrer Gruppenzugehörigkeit, dann können Unterschiede in der Beantwortung der Indikatoren zwischen Befragten verschiedener Staaten als Unterschiede in den Vorurteilen interpretiert werden. Formal lässt sich dies als konditionale

11 In den vergangenen Jahren wurden auch latente Klassenanalysen (z. B. Eid et al. 2003) und multiple Korrespondenzanalysen eingesetzt (Blasius und Nießen 2006).

12 Aus Gründen der Modellidentifikation wird der latente Mittelwert einer Referenzgruppe dabei auf Null fixiert, d. h. die geschätzten Parameter geben die Abweichungen von der Referenzgruppe an.

Wahrscheinlichkeit einer bestimmten Ausprägung x eines Indikators X , gegeben den Faktorwert ξ^{13} , darstellen (j bezeichnet den Laufindex über die Variablen, i den Laufindex über die Personen und k den Laufindex über die Gruppen). Die konditionale Wahrscheinlichkeit ist dann unabhängig von der Gruppenzugehörigkeit k einer Person i , sie hängt ausschließlich vom Faktorwert ab (vgl. im folgenden Millsap & Yun-Tein 2004, 483 ff.).

$$P_k(X_{ijk} = x | \xi_{ik}) = P(X_{ijk} = x | \xi_{ik})$$

Wann liegt Messinvarianz auf diesem Niveau vor? Weil sich das Faktorenmodell bei kategorialen Indikatoren auf die zugrundeliegenden Variablen X_{ijk}^* bezieht¹⁴, Meßinvarianz aber für die beobachteten Indikatoren X_{ijk} definiert ist, müssen die *Schwellenwerte* für die Indikatoren über alle Gruppen gleich sein: $v_{jk} = v_j$.

Damit die Indikatoren in allen Staaten dasselbe messen, müssen die *Faktorladungen* (die sich hier auf die zugrundeliegenden Variablen beziehen) über alle Gruppen identisch sein: $\lambda_{jk} = \lambda_j$. Dies beinhaltet die Forderung einer gleichen Faktorenstruktur. Im Beispiel müssten demnach die vier Indikatoren in allen Staaten eindimensional messen. Sind die Faktorladungen nicht identisch, dann erfassen die (zugrundeliegenden) Variablen nicht in allen Staaten in gleicher Weise die Zieldimension (Vorurteile). Die Validität unterscheidet sich dann nach Staaten. Zudem müssen die Konstanten des Messmodells (Fußnote 14) identisch sein: $\tau_{jk} = \tau_j$. Bei nicht-identischen Konstanten des Messmodells wäre die von den Vorurteilen unabhängige Ausprägung auf X_{ijk}^* für die Staaten verschieden, weshalb Differenzen in den X_{ijk}^* -Werten nicht als Unterschiede in den Vorurteilen interpretiert werden dürften. Schließlich wird gefordert, dass die *Residualvarianzen* (Messfehler) der Indikatoren über alle Gruppen gleich sind: $\delta_{jk} = \delta_j$. Gleiche Reliabilitäten werden aber in der Literatur nicht einheitlich als notwendige Voraussetzung für den Vergleich von Mittelwerten auf der latenten Ebene angesehen. Um eine inhaltliche Interpretation von Unterschieden auf der Gruppenebene vornehmen zu können, ist es *nicht* notwendig, dass die Varianzen des Faktors in den Gruppen (bei mehreren Faktoren auch die Kovarianzen zwischen den Faktoren) identisch sind.

Zur Schätzung wird Mplus verwandt (Muthén und Muthén 2006). Mplus bietet sich an, weil direkt die gemeinsamen Schätzwerte über die imputierten Datensätze berechnet werden. Zur Modellidentifikation werden in Mplus die Konstanten der Messgleichung auf null fixiert, d. h. diese können nicht frei geschätzt werden.

In welcher Reihenfolge die Restriktionen getestet werden, hängt von der Fragestellung ab (vgl. Reinecke 2005). Im folgenden wird als Ausgangsmodell ein Basismodell gewählt, in dem Faktorladungen und Schwellenwerte über alle Gruppen frei variieren können.¹⁵ Mit diesem Basismodell wird damit geprüft, ob die unterstellte Faktorenstruktur (hier: Eindimensionalität) gerechtfertigt werden kann. Dieses Modell ist nur dann identifiziert, wenn die Residualvarianzen der Indikatoren auf eins und die Mittelwerte des Faktors auf null fixiert werden.¹⁶ In einem zweiten Schritt wird ein Modell

13 Bzw. bei einem mehrfaktoriellen Modell: der gemeinsamen Faktorwerte.

14 Im kongenerischen Fall: $X_{ijk}^* = \tau_{jk} + \lambda_{jk}\xi_{ik} + u_{ijk}$.

15 Als Schätzverfahren wurde WLSMV verwandt (Muthén und Muthén 2006, 426).

16 In einem Modell mit dichotomen Indikatoren lassen sich Schwellenwerte, Faktorladungen und Residualvarianzen nicht gleichzeitig frei schätzen, weil das Modell dann nicht identifiziert ist (vgl. dazu detailliert Millsap und Yun-Tein, 507 f.). Zur Modellidentifikation bei dichotomen Indikatoren wurde alternativ zum

geschätzt, in dem Schwellenwerte *und* Faktorladungen über alle Gruppen restringiert sind. Schwellenwerte und Faktorladungen werden in einem Schritt über die Gruppen gleichgesetzt, weil die Itemcharakteristik der dichotomen Indikatoren von Schwellenwerten und Faktorladungen gleichzeitig beeinflusst wird (Muthén und Muthén 2006, 345). Sofern dieses Modell angemessen ist, können in einem weiteren Schritt zusätzlich die Residualvarianzen fixiert werden. Damit unterscheidet sich die Reihenfolge der eingeführten Restriktionen von einem metrischen Messmodell. Die Modellqualität wird anhand des Comparative Fit Index und des RMSEA beurteilt. (Robuste) χ^2 -Differenzentests zwischen hierarchisch geschachtelten Modellen sind nicht über alle imputierten Datensätze möglich, weshalb diese nicht zum Modellvergleich herangezogen werden können. Führt man Gruppenvergleiche über mehr als zwei Gruppen durch, dann ist es nicht sinnvoll, eine Restriktion (hier gleiche Faktorladungen und Schwellenwerte) direkt über alle Gruppen aufzugeben. Sofern das Modelle mit restringierten Faktorladungen und Schwellenwerten keine gute Modellanpassung aufweist, werden daher Faktorladungen und Schwellenwerte für einzelne Länder freigesetzt. Als Hinweis für eine schlechte Modellanpassung in einzelnen Ländern dienen deren χ^2 -Beiträge.¹⁷

Ergebnisse

Die Ergebnisse des Modellvergleichs finden sich in Tabelle 4. Das Basismodell, das lediglich die Struktur der Faktoren (Eindimensionalität) vorgibt, weist eine gute Modellanpassung auf. Das Modell mit gleichen Schwellenwerten und Faktorenladungen erzielt nach den Kriterien von Browne und Cudeck (1993) jedoch keine akzeptable Modellanpassung. Ursächlich dafür ist vor allem eine schlechte Modellanpassung in Großbritannien, Dänemark, Griechenland, Irland und Frankreich. Die Freisetzung von Faktorladungen und Schwellenwerten für einzelne Indikatoren dieser vier Staaten führt zu einer deutlichen Verbesserung der Modellanpassung, die mit einem RMSEA von .05 als gut eingestuft werden kann. Trotz der freigesetzten Parameter ist die Modellanpassung für Großbritannien jedoch deutlich schlechter als in den anderen Staaten.¹⁸

In der Analyse von Semyonov et al. (2006) werden damit Unterschiede im Ausmaß von Vorurteilen zwischen Staaten modelliert, die in Großbritannien, Griechenland, Irland und Frankreich (auch) das Resultat einer fehlenden Invarianz der Indikatoren sein können. Anders ausgedrückt: Unterschiede in der Zustimmung bzw. Ablehnung der Indikatoren zwischen den untersuchten Staaten spiegelt für diese Staaten nicht ausschließlich Unterschiede in den Vorurteilen wider.

hier gewählten Vorgehen vorgeschlagen, die Schwellenwerte über die Gruppen gleich zu setzen und die Residualvarianzen der Variablen, die zur Skalierung der latenten Variablen dienen, auf eins zu fixieren (vgl. Millsap und Yun-Tein, 508).

17 Diese - und die Modifikationsindizes - erhält man nur in getrennten Analysen der einzelnen imputierten Datensätze.

18 Ohne Großbritannien nimmt der RMSEA für Modell 2 den Wert .040 und der CFI den Wert .995 an.

Tabelle 4: Messinvarianz der Vorurteilsindikatoren

| | Arbeitslos. | Sozialsys. | Unsicher. | Zahl | CFI | RMSEA |
|---|-------------|------------|-----------|------|------|-------|
| Basismodell | | | | | | |
| ν, λ frei | | | | | .997 | .043 |
| Modell 1: vollständige Invarianz | | | | | | |
| ν, λ invariant | | | | | .962 | .108 |
| Modell 2: partielle Invarianz | | | | | | |
| ν und λ partiell frei für | | | | | .992 | .050 |
| Großbritannien | | | frei | frei | | |
| Dänemark | | | frei | | | |
| Griechenland | | frei | | | | |
| Irland | | frei | | | | |
| Frankreich | | frei | | | | |

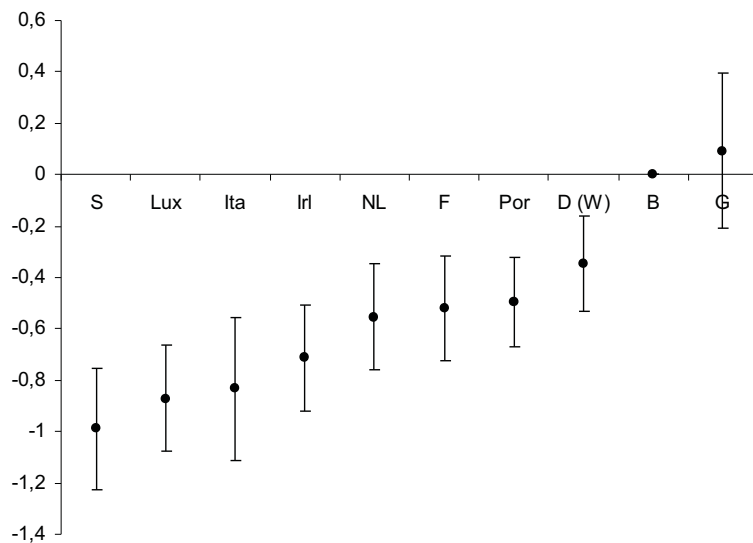
6372 Gemeinsame Schätzung über die 5 vervollständigten Datensätze.

Problematisch ist vor allem der Indikator „Sozialsystem“. Ohne diesen Indikator kann für zehn der zwölf Staaten (alle außer Großbritannien und Dänemark) von einer Messinvarianz ausgegangen werden, die eine inhaltliche Interpretation von Unterschieden auf der latenten Ebene ermöglicht. Für diese zehn Staaten wurde Modell 2 mit den Indikatoren Arbeitslosigkeit, Unsicherheit und Zahl geschätzt. Dieses Modell erzielt einen RMSEA von .042 und einen CFI von .995.¹⁹ Die Mittelwerte in den Vorurteilen für die einzelnen Staaten lassen sich aus Gründen der Modellidentifikation nur als Abweichung von einer Referenzgruppe bestimmen. Für die Referenzgruppe - hier Belgien - wird der latente Mittelwert auf null fixiert. Wie man sieht, sind die Vorurteile in Griechenland, Belgien und Westdeutschland am höchsten und in Italien, Luxemburg und Spanien am niedrigsten ausgeprägt. Angegeben sind auch die 95%igen Konfidenzintervalle: die Vorurteile unterscheiden sich in einer ganzen Reihe von Staaten nicht signifikant. Signifikant höher als in fast allen anderen Staaten fallen die Vorurteile in Griechenland aus.

Vergleicht man die Ergebnisse mit denen von Semyonov et al. (2006), siehe Abbildung 2, dann fällt vor allem die unterschiedliche Position der französischen und irischen Befragten auf. Deren Vorurteile werden durch Verwendung des Indikators Sozialsystem bei Semyonov et al. (2006) überschätzt. Französische und irische Befragte stimmen dem Indikator Sozialsystem bei gleichen Einstellungen eher zu als Befragte anderer Staaten, weshalb die Zustimmung zu diesem Indikator nicht in gleichem Maße wie in anderen Staaten Vorurteile widerspiegelt.

In den meisten Studien erfolgt die Skalenbildung auf Basis der manifesten Indikatoren. Im Hinblick auf die drei Vorurteilsindikatoren würde eine Summenskala die Zahl der Zustimmungen zu diesen Indikatoren angeben, die von null - Zustimmung zu keinem der drei Items - bis drei - Zustimmung zu allen drei Items - reicht. Zu erkennen sind deutliche Unterschiede in der Verteilung der Vorurteile. Die französischen Befragten sind hinsichtlich ihrer Vorurteile in zwei polarisierende „Lager“ gespal-

¹⁹ Schwellenwerte: Arbeitslosigkeit -.90, Unsicherheit -.57 und Zahl -.22; Faktorladungen: Arbeitslosigkeit 1 (zur Skalierung des Faktors fixiert), Unsicherheit .944 und Zahl .983. Schwellenwerte und Faktorladungen sind signifikant.

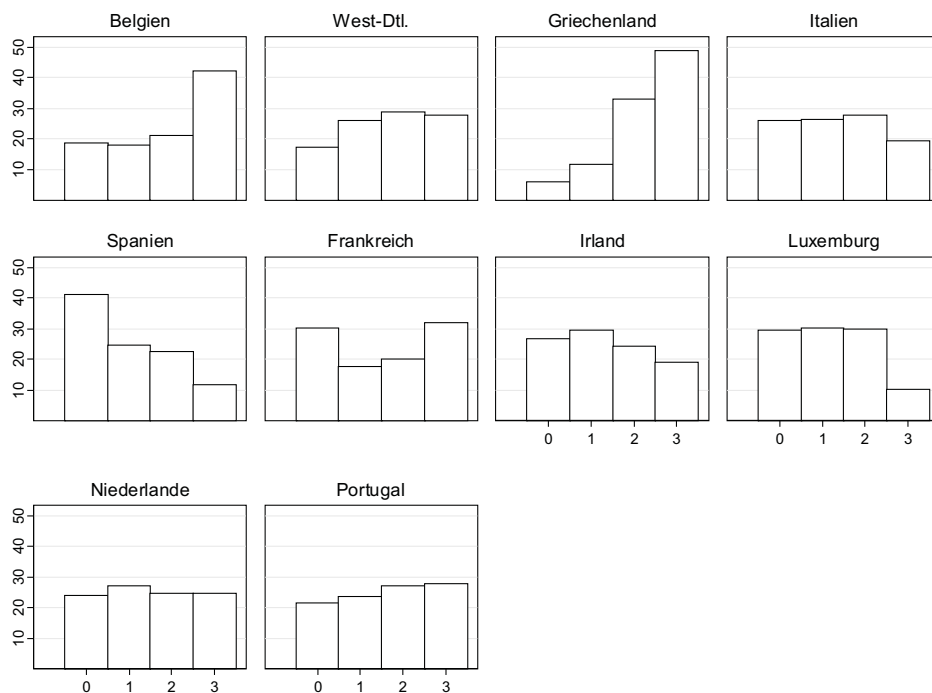


Gemeinsame Schätzung über die 5 vervollständigten Datensätze. RMSEA .042, CFI .995

Abbildung 3: Mittelwertdifferenzen (und 95%ige Konfidenzintervalle) in den Vorurteilen

ten, nämlich in Befragte, die entweder alle Aussagen ablehnen oder allen Aussagen zustimmen. Bemerkenswert ist zudem, dass nahezu die Hälfte aller griechischen und mehr als 40% der belgischen Befragten allen drei Aussagen zustimmt, wobei die Anteile der griechischen Befragten ohne bzw. mit geringen Vorurteilen vernachlässigt werden können. In Spanien ist das Bild genau umgekehrt: ca. 40% der Befragten stimmen hier keiner der Aussagen zu.

Ca. 9% der Unterschiede in den Vorurteilen ($p = 0,09$, $s.e.=0,04$) entfallen auf Unterschiede zwischen Befragten verschiedener Staaten, 91% auf Unterschiede zwischen Befragten innerhalb von Staaten. Die Unterschiede zwischen Staaten sind auf dem 5%-Niveau statistisch signifikant. Der Anteil der Varianz, der auf die Staaten entfällt, resultiert dabei wesentlich aus den Unterschieden im Niveau der Vorurteile zwischen Griechenland und den restlichen Staaten (ohne Griechenland: $p = 0,05$, $s.e.=0,02$).



Gemeinsame Schätzung über die 5 vervollständigten Datensätze

Abbildung 4: Skala fremdenfeindlicher Vorurteile (Zahl zustimmender Angaben in Prozent)

Konsequenzen für die vergleichende Forschung

In der komparativen Vorurteilsforschung werden Unterschiede im Ausmaß von Vorurteilen zwischen Staaten analysiert und durch eine Reihe von Prädiktoren auf Ebene der Staaten (wie die Höhe der Zuwanderung) zu erklären versucht. Eine inhaltliche Interpretation von Unterschieden in den Skalenwerten von Befragten verschiedener Staaten ist jedoch nur möglich, wenn die zur Bildung der Vorurteilsskalen verwandten Indikatoren in allen Staaten dasselbe messen. Um so erstaunlicher ist es, dass die Überprüfung der Messinvarianz in komparativen Analysen von Vorurteilen kaum eine Rolle spielt - z. B. dann, wenn die Dimensionalität der verwendeten Indikatoren nicht für jeden Staat getrennt überprüft wird. Dabei stehen etablierte Methoden wie multiple Gruppenvergleiche konfirmatorischer Faktorenanalysen zur Überprüfung der Messinvarianz in verschiedenen Gruppen zur Verfügung, die in einer Reihe von Programmen (LISREL, Amos, EQS, Mplus) implementiert sind. Eine inhaltliche Interpretation von Unterschieden zwischen Gruppen auf der latenten Ebene als Unterschiede im Ausmaß der Vorurteile ist bei dichotomen oder ordinalen Indikatoren möglich, wenn die Parameter des Faktorenmodells und die Schwellenwerte für die verglichenen Staaten identisch sind. Für die von Semyonov et al. (2006) verwandten vier Indikatoren zur Messung von Vorurteilen kann auf diesem Niveau in sieben der zwölf Staaten Messinvarianz festgestellt werden. Die Zahl der Staaten, für die Vergleiche im Ausmaß der Vorurteile mit diesen vier Indikatoren möglich sind, ist

daher begrenzt. Eine alternative Strategie besteht darin, Indikatoren auszuschließen, die nicht in allen Staaten dasselbe messen, um den Vergleich von Vorurteilen in einer größeren Zahl von Staaten zu ermöglichen.

Mehrebenenanalysen, die zur Überprüfung der Thesen zur Erklärung von Unterschieden im Ausmaß von Vorurteilen zwischen Staaten verwandt werden, setzen (deutlich) mehr als zehn Staaten voraus (Snijders und Bosker 1999, 140, 154). Die Überprüfung der Messinvarianz der Indikatoren hat hier dazu geführt, dass die (onehin knappe) Zahl der Einheiten auf Ebene 2 (Staaten) nicht mehr zur Anwendung mehrebenenanalytischer Verfahren ausreicht. Berücksichtigt werden muss allerdings, dass nur ein geringerer Teil der Unterschiede in den Vorurteilen von Befragten auf Gruppenunterschiede zurückzuführen sind. In der vorliegenden Analyse waren es knapp 9%, in anderen Studien variiert der Anteil der Varianz auf der Ebene der Staaten zwischen 5 und 15% (Weins 2004, 34). Die Unterschiede zwischen Befragten innerhalb von Staaten sind damit deutlich größer (zwischen 85 und 95 Prozent) als die Unterschiede zwischen Befragten verschiedener Staaten. Eingangs wurden die Ergebnisse von Quillian (1995) berichtet, wonach ein erheblicher Teil der Unterschiede im Ausmaß von Vorurteilen zwischen Staaten (ca. 70%) durch die Größe der Minderheit und die makro-ökonomische Lage erklärt werden. Weil Gruppenunterschiede in seiner Analyse ca. 10% der Unterschiede in den Vorurteilen ausmachen,²⁰ erklären die Kontextmerkmale damit „nur“ ca. 7% (nämlich 70% von 10%) der Gesamtvarianz in den Vorurteilen. Das heißt natürlich nicht, dass auf eine mehrebenenanalytische Modellierung nur aufgrund geringerer Varianzanteile auf Ebene der Staaten verzichtet werden sollte. Ausschlaggebend sollten hier theoretische Argumente sein. Die Bedeutung der Merkmale auf Ebene der Staaten für die Entstehung von Vorurteilen relativiert sich jedoch erheblich. Für komparative Analysen heißt dies, dass die theoretische Fundierung und Überprüfung des Einflusses individueller Merkmale nicht vernachlässigt werden sollte. Weil sich dieser Beitrag mit der internationalen Vergleichbarkeit von Messinstrumenten beschäftigt, sei abschließend bemerkt, dass der in den zitierten Studien verwendete Indikator für die Größe der zugewanderten Minderheit, nämlich der Anteil der Ausländer (bzw. der Anteil der Nicht-EG/EU-Ausländer) für international vergleichende Studien kein geeignetes Messinstrument ist. Ausländeranteile hängen wesentlich vom Staatsangehörigkeitsrecht ab, und die gesetzlichen Regelungen zur Staatsangehörigkeit variieren bekanntermaßen selbst zwischen den europäischen Staaten deutlich. Es erstaunt nicht, dass der Anteil der Nicht-EU-Ausländer in der Analyse von Semyonov et al. (2006) ausgerechnet für Frankreich gering ist.²¹ Die (zitierten) substanzwissenschaftlichen Ergebnisse zum Einfluss des Ausländeranteils auf die Entstehung von Vorurteilen müssen daher mit Vorsicht interpretiert werden. Ein international besser vergleichbarer Indikator für die Größe der (zugewanderten) Minderheit wäre der nicht im (Befragungs-)Staat geborene Bevölkerungsanteil.

Full-score comparability, definiert als Abwesenheit jeglicher Verzerrung, die den Vergleich von Skalenwerten Befragter verschiedener Gruppen beeinträchtigen könnte (Van de Vijver 2003), ist nicht direkt nachweisbar und kann nur durch den Ausschluss möglicher Fehlerquellen angestrebt werden. In diesem Beitrag wurden Beeinträchtigungen in der internationalen Vergleichbarkeit von

20 Quillian hat in seinem Beitrag die Varianzanteile auf Ebene der Staaten und der Individuen nicht veröffentlicht, sendet diese aber auf Nachfrage zu.

21 Ein Staat mit restriktiveren Regelungen zum Erwerb des Staatsangehörigkeitsrechtes für Zuwanderer und deren Nachfahren hat c. p. bei gleich hoher Zuwanderung einen höheren Ausländeranteil. Der Anteil der Ausländer aus Nicht-EU-Staaten in Frankreich auch deshalb so gering, weil es sich bei Zuwanderern aus Nicht-EU-Staaten vorwiegend um Zuwanderer aus den ehemaligen (nordafrikanischen) Kolonien handelt, die - sofern sie vor der Unabhängigkeit Algeriens geboren wurden, französische Staatsbürger sind.

Messinstrumenten aufgrund von Ausfällen (item-nonresponse) thematisiert und durch die mehrfache Ergänzung fehlender Werte eingegrenzt. Das hier eingesetzte Instrumentarium ist flexibel genug, um weitere Fehlerquellen, die den internationalen Vergleich von Vorurteilen beeinträchtigen könnten (z. B. Aquieszenz, soziale Erwünschtheit) direkt zu kontrollieren, indem Instrumente zur Messung solcher Fehlerquellen als unabhängige (latente oder manifeste) Variablen modelliert werden (vgl. z. B. Billiet 2003). Die Grenzen einer solchen Kontrolle werden bei sekundäranalytischen Auswertungen durch die Datenlage (z. B. Skalen zur Messung sozialer Erwünschtheit) vorgegeben.

Literatur

- Allison, Paul D. (2002). *Missing Data*. Thousand Oaks: Sage.
- Billiet, Jaak (2003). Cross-Cultural Equivalence with Structural Equation Modeling, 247-263 in *Cross Cultural Survey Methods*, hrsg. von Janet A. Harkness, Fons van de Vijver und Peter Mohler Hoboken, NJ: Wiley.
- Blalock, Hubert M. (1967). *Toward a Theory of Minority-Group Relations*. New York: Wiley.
- Blasius, Jörg und Victor Thiessen (2006). Assessing Data Quality and Construct Comparability in Cross-National Surveys. *European Sociological Review* 22, 229-242.
- Bollen, Kenneth A. (1989). *Structural Equations With Latent Variables*. Wiley.
- Braun, Michael (2006). *Funktionale Äquivalenz in interkulturell vergleichenden Umfragen. Mythos und Realität*. Mannheim: ZUMA.
- Coenders, Marcel (2001). *Nationalism and Ethnic Exclusionism in a Comparative Perspective*. Nijmegen: ICS Dissertation.
- Couper, Mick P. und Edith de Leeuw (2003). Nonresponse in Cross-Cultural and Cross-National Surveys. 157-177 in *Cross Cultural Survey Methods*, hrsg. von Janet A. Harkness, Fons van de Vijver und Peter Mohler Hoboken, NJ: Wiley.
- Eid, Michael, Langeheine, Rolf & Diener, Ed (2003). Comparing Typological Structures Across Cultures by Latent Class Analysis: A primer. *Journal of Cross-Cultural Psychology* 34, 195-210.
- Engelhardt, Henriette (1999). Lineare Regression mit Selektion: Möglichkeiten und Grenzen der Heckman-Korrektur. *KZfSS* 51, 706-723.
- Francis, Emerich K. (1965). *Ethnos und Demos*. Berlin: Duncker und Humblot.
- Harkness, Janet A., Vijver, Fons van de und Peter Mohler (2003). *Cross Cultural Survey Methods*. Hoboken, NJ: Wiley.
- INRA (2000). *Eurobarometer 53. April-May 2000. Codebook*.
- Kunovich, Robert M. (2001). *Group Threat and Attitudes Toward Immigrants*. Dissertation. Ohio-State-University .
- Little, Roderick J. A. und Donald B. Rubin (2002). *Statistical Analysis With Missing Data*. 2. Auflage. New York: Wiley.
- Millsap, Roger E. und Jenn Yun-Tein (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research* 39, 479-515.
- Muthén, Linda K. und Bengt O. Muthén (2006). *Mplus. Statistical Analysis with Latent Variables. User's Guide*. 4. Auflage. Los Angeles.

-
- Olsson, Ulf (1979). Maximum Likelihood Estimation of the Polychoric Correlation. *Psychometrika* 54, 557-585.
- Pan, Christoph und Beate Pfeil (2000). *Die Volksgruppen in Europa: ein Handbuch*. Wien: Braumüller.
- Quillian, Lincoln (1995). Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe. *American Sociological Review* 60, 586-611.
- Reinecke, Jost (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg.
- Royston, Patrick (2005). Multiple Imputation of Missing Values: Update. *Stata Journal* 5, 188-201.
- Royston, Patrick (2004). Multiple Imputation of Missing Values. *Stata Journal* 4, 227-241.
- Scheepers, Peer, Mérove Gijsberts und Marcel Coenders (2002). Ethnic Exclusionism in European Countries. *European Sociological Review* 18, 17-34.
- Schräpler, Jörg-Peter (2004). Respondent Behavior in Panel Studies. A Case Study for Income Non-response by Means of the German Socio-Economic Panel (SOEP). *Sociological Methods and Research* 33, 118-156.
- Semyonov, Moshe, Rebeca Raijman und Anastasia Gorodzeisky (2006). The Rise of Anti-foreigner Sentiment in European Societies, 1988-2000. *American Sociological Review* 71, 426-449.
- Ter Wal, Jessika und Maykel Verkuyten (2000). Introduction: Racism in a Comparative Perspective, 1-21 in: *Comparative Perspectives on Racism*. Aldershot u. a.: Ashgate.
- Vijver, Fons van de und Kwok Leung (1997). *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks u. a.: Sage
- Vijver, Fons van de (2003). Bias and Equivalence: Cross-Cultural Perspectives, 143-155 in *Cross Cultural Survey Methods*, hrsg. von Janet A. Harkness, Fons van de Vijver und Peter Mohler. Hoboken, NJ: Wiley.
- Weins, Cornelia und Jost Reinecke (2007). Delinquenzverläufe im Jugendalter: Eine methodologische Analyse zur Auswirkung von fehlenden Werten im Längsschnitt. *Monatsschrift für Kriminologie und Strafrechtsreform* 90, 418-437.
- Westle, Bettina (1999). *Kollektive Identität im vereinten Deutschland. Nation und Demokratie in der Wahrnehmung der Deutschen*. Opladen: Leske und Budrich.
- Wilkes, Rima, Neil Guppy und Lily Farris (2007). Right-Wing Parties and Anti-Foreigner Sentiment in Europe. Comment on Semyonov, Raijman, and Gorodzeisky, ASR, June 2006. *American Sociological Review* 72, 831-840.
- Zick, Andreas (1997). *Vorurteile und Rassismus. Eine sozialpsychologische Analyse*. Münster: Waxmann.

Zur Person

Dr. Cornelia Weins,
wiss. Assistentin an der Abteilung Soziologie der Universität Trier.

Forschungs- und Lehrschwerpunkte:
Sozialstrukturanalyse, politische Soziologie und Methodenlehre.

Aktuelle Publikation:
Delinquenzverläufe von Jugendlichen (gemeinsam mit Jost Reinecke). Monatsschrift für Kriminologie und Strafrechtsreform 90 (2007), 418-437.

Kontakt:
Dr. Cornelia Weins
Abteilung Soziologie
Universität Trier
54286 Trier

weins@uni-trier.de